BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

EXHIBIT B

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL LABEL NUMBER EM299801549US
DATE OF DEPOSIT: February 4, 1998

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER OF PATENTS, BOX PATENT APPLICATION, WASHINGTON, D.C. 20231

Tae Kim
(TYPED OR PRINTED NAME OF PERSON MAILING PAPER)

(SIGNATURE OF PERSON MAILING PAPER OR FEE)

APPLICATION

for

UNITED STATES LETTERS PATENT

on

Microarrays and Uses Therefor

by

James Hoeffler

and

Joseph Fernandez

Sheets of Drawings: none

Docket No.: INVIT1100

Attorneys

Gray Cary Ware & Freidenrich
4365 Executive Drive, Suite 1600
San Diego, California 92121-2189

MICROARRAYS AND USES THEREFOR

Field of the Invention

The invention disclosed herein relates to new methods of using microarray technologies. The methods are useful for identifying and characterizing specific antibodies as well as the characterization of different tissues or cells by nucleic acid or protein analysis.

Background of the Invention

Recent breakthroughs in nucleic acid sequencing technology have made possible the sequencing of entire genomes from a variety of organisms, including humans. The potential benefits of a complete genome sequence are many, ranging from applications in medicine to a greater understanding of evolutionary processes. These benefits cannot be fully realized, however, without an understanding of how and where these newly sequenced genes function.

Traditionally, functional understanding started with recognizing an activity, isolating a protein associated with that activity, then isolating the gene, or genes, encoding that protein. The isolated protein was also used to generate antibody reagents. Specific antibodies and fragments of the isolated gene were both employed to study tissue expression and function.

Several methods have been used to study protein expression patterns including *in situ* hybridization studies of tissue sections and Northern blots. These methods are both time consuming and require relatively large amounts of material to perform successfully.

Antibodies that bind to specific antigens have been produced by a variety of methods including immunization of animals, fusion of mammalian spleen cells to immortalized cells to produce hybridomas, random peptide generation using phage or bacterial display and constrained peptide libraries. Regardless of how the desired antibody is generated the methods currently available to identify one with a particular binding specificity are generally laborious and incapable of the simultaneous testing of large numbers of unknowns.

One method involves binding the antigen to a porous membrane, such as nitrocellulose, contacting the membrane with a source of test antibodies, then determining whether or not any of the test antibodies has bound to the antigen. This method only allows the testing of one source of test antibodies per piece of porous membrane, making the method both inconvenient and wasteful of materials.

Antibody/antigen reactions can also be evaluated in plastic plates, such as 96-well microtiter plates, using methods similar to those described above. This method is likewise limited in the number of samples that can be tested in any one assay, thus requiring many assays to fully evaluate a large number of antibody unknowns.

Recently new technologies have arisen that allow the creation of microarrays containing thousands or millions

of different elements. Such array technology has been applied mainly to forming arrays of individual nucleic acids (see, for example, Marshall and Hodgson, Nature Biotech. 16:27-31, 1998; Ramsay, Nature Biotech. 16:40-44, 1998).

A need exists for methods of assessing gene function. Methods are needed to simply and rapidly screen very large numbers of uncharacterized antibodies for those specific for a given antigen as well as for the characterization of tissues and cells by nucleic acid and/or protein analysis. The invention described herein addresses that need.

Brief Description of the Invention

The invention disclosed herein comprises methods of using microarrays to simplify analysis and characterization of genes and their function. In one aspect of the invention the methods are used to identify and characterize antibodies having binding affinity for a specific target antigen. This method comprises contacting an array of uncharacterized antibodies bound to a solid surface with at least one target antigen and identifying the antibodies to which the target antigen binds. The method can be performed under a variety of conditions to identify antibodies with a range of binding affinities.

A second aspect of the invention comprises a method of determining gene expression at the protein level comprising contacting an array of characterized or

uncharacterized antibodies on a solid surface with one or more proteins and identifying the antibodies to which said protein(s) binds. This method can be further used to compare the protein expression in two different populations of cells, such as normal cells and cancer cells or resting cells and stimulated cells.

A further aspect of the invention comprises a method of determining gene expression at the protein level comprising contacting a microarray of nucleic acid samples derived from a variety of different sources with one or more nucleic acid probes then identifying the sample or samples to which the probe binds.

Detailed Description of the Invention

The present invention discloses methods of using microarrays to simplify analysis and characterization of genes and their function. In a first aspect of the invention the methods are used for identifying and characterizing antibodies having binding specificity to a particular antigen or set of antigens. This method utilizes microarray technology to create ordered matrices of very large numbers of uncharacterized antibodies which can then be contacted with antigen under a variety of conditions. The method is rapid and simple to perform and is applicable to the simultaneous screening of very large numbers of antibodies.

Briefly, uncharacterized, isolated antibodies are bound to a solid surface in an array format consisting of discrete spots whose spatial location can be easily

identified. Each location represents an antibody from a known source, such as a particular hybridoma growing in a well in a 96-well microtiter plate. The space between the antibody spots is treated to minimize non-specific binding to the solid support. The arrayed antibodies are then contacted with an antigen, or a set of antigens, for which specific antibodies are sought. The antigen solution is left in contact with the array for an amount of time sufficient to allow antigen:antibody complexes to form, then the unbound antigen is washed awayunder suitable conditions. Bound antigen is detected at a particular antibody spot using one of a variety of detection methods, thus identifying the source of an antibody specific for the particular antigen.

The term "antibody" is used herein in the broadest sense and specifically includes intact monoclonal antibodies, polyclonal antibodies, multispecific antibodies (*e.g.* bispecific antibodies) formed from at least two intact antibodies, and antibody fragments, including single chain antibodies, so long as they exhibit the desired binding properties as described herein.

Various procedures well-known in the art may be used for the production of polyclonal antibodies to an epitope or antigen of interest. A host animal of any of a number of species, such as rabbits, goats, sheep, horse, cow, mice, rats, etc. is immunized by injection with an antigenic preparation which may be derived from cells or microorganisms, or may be recombinantly or synthetically produced. Various adjuvants well known in the art may be used to enhance the production of antibodies by the immunized host, for example, Freund's adjuvant (complete

and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, liposomes, potentially useful human adjuvants such as BCG (Bacille Calmette-Guerin) and *Propionibacterium acanes*, and the like.

The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. Preferred antibodies are mAbs, which may be of any immunoglobulin class including IgG, IgM, IgE, IgA, and any subclass or isotype thereof.

In addition to their specificity, monoclonal antibodies are advantageous in that they are synthesized by hybridoma culture, uncontaminated by other immunoglobulins. The modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.*, *Nature*, 256:495 (1975),

or may be made by recombinant DNA methods (see, *e.g.*, U.S. Patent No. 4,816,567, incorporated by reference herein). The "monoclonal antibodies" may also be isolated from phage antibody libraries using the techniques described in Clackson *et al.*, *Nature*, 352:624-628 (1991) and Marks *et al.*, *J. Mol. Biol.*, 222:581-597 (1991), for example.

The monoclonal antibodies contemplated for use herein specifically include "chimeric" antibodies (immunoglobulins) in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity (U.S. Patent No. 4,816,567; Morrison *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:6851-6855 (1984)).

"Humanized" forms of non-human (*e.g.*, murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementarity-determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity,

affinity, and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and maximize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see Jones *et al.*, *Nature*, 321:522-525 (1986); Reichmann *et al.*, *Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992). The humanized antibody includes a PRIMATIZED™ antibody wherein the antigen-binding region of the antibody is derived from an antibody produced by immunizing macaque monkeys with the antigen of interest.

"Antibody fragments" comprise a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', $F(ab')_2$, and Fv fragments; diabodies; linear antibodies (Zapata *et al. Protein Eng.* 8(10):1057-1062 (1995)); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

Particularly preferred in the practice of the invention are single-chain antibodies. "Single-chain" or "sFv" antibodies are antibody fragments comprising the V_H and V_L domains of an antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the sFv to form the desired structure for antigen binding. For a review of sFvs see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

Large quantities of single chain antibodies with uncharacterized randomized binding specificity can be produced using a number of methodologies known in the art. Random peptide libraries can be created in filamentous phage particles (Daniels and Lane, Methods 9(3):494-507, 1996; Reichmann and Weill, Biochemistry 32(34):8848-8855; Rader and Barbas, Curr Opin Biotechnol 9(4):503-508, 1997; Iba and Kurosawa, Immunol Cell Biol 75(2):217-221, 1997), for example, or similarly in yeast, bacteria, and the like. Other methods for creating random libraries of sFvs include various solid state synthesis methods.

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) in the same polypeptide chain (V_H - V_L). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for

example, EP 404,097; WO 93/11161; and Hollinger *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:6444-6448 (1993).

The antibodies employed in the invention are generally isolated prior to creating a microarray. An "isolated" molecule, whether an antibody, antigen or nucleic acid, is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the molecule, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, a protein will be purified (1) to greater than 95% by weight of protein as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated protein includes the protein *in situ* within recombinant cells since at least one component of the protein's natural environment will not be present. Ordinarily, however, isolated protein will be prepared by at least one purification step.

By "isolated" in reference to nucleic acid is meant a polymer of 14, 17, 21 or more contiguous nucleotides, including DNA or RNA that is isolated from a natural source or that is synthesized. The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally

occurring sequence has been removed from its normal cellular (i.e., chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide sequence present, but that it is essentially free (about 90 - 95% pure at least) of non-nucleotide material naturally associated with it and thus is meant to be distinguished from isolated chromosomes.

One particularly useful method of isolating antibodies, such as single chain antibodies from a cell extract, is affinity purification using *his* resin or affinity tags. By using an amount of resin with binding sites sufficient for only a small portion of the antibody present in the unpurified mixture, the process of isolation can be used to simultaneously normalize yield and isolate the antibody. For example, although each sample will contain a different and unknown amount of antibody protein, the samples can be contacted with an amount of resin whose maximum binding capacity is 10 mgs. Thus any antibody greater than this amount will pass through the resin unbound. The maximum bound amount can then be eluted from the resin. Resins suitable for antibody purification are well known in the art, for example, protein A sepharose.

Methods for creating microarrays are known in the art. One method is described in Shalon and Brown (WO 95/35505, published 12/28/95) which is incorporated herein by reference in its entirety. The method and apparatus described in Shalon and Brown can create an array of up to one million spots per cubic centimeter on a glass slide using a volume of 0.01 to 100 nl per spot.

In the present invention, each spot can contain one or more than one distinct uncharacterized antibody.

Other methods of creating arrays are known in the art, including photolithographic printing (Pease, et al, PNAS 91(11):5022-5026, 1994) and in situ synthesis. While known in situ synthesis methods are less useful for synthesizing polypeptides long enough to be antibodies, they can be used to make polypeptides up to 50 amino acids in length, which can serve as binding proteins as described below.

The microarrays can be created on a variety of solid surfaces such as plastics (eg. polycarbonate), complex carbohydrates (eg. agarose and sepharose), acrylic resins(eg. polyacrylamide and latex beads), and nitrocellulose. Preferred solid support materials include glass slides and silicon wafers.

Methods for covalent attachment of antibodies to a solid support are known in the art. Examples of such methods are found in Bhatia, et al, Anal. Biochem. 178(2):408-413, 1989; Ahluwalia, et al, Biosens. Bioelectron. 7(3):207-214, 1992; Jonsson, et al, Biochem. J. 227(2):373-378, 1985; and Freij-Larsson, et al, Biomaterials 17(22):2199-2207, 1996, all of which are incorporated by reference herein in their entirety. Proteins may additionally be attached to a solid support using methods described in the examples below.

Methods of reducing non-specfic binding to a solid surface are well known in the art and include washing the arrayed solid surface with bovine serum albumin (BSA), reconstituted non-fat milk, salmon sperm DNA, porcine

heparin and the like (see Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995).

The arrays used to identify antigen-specific antibodies are contacted with a solution containing one or more known antigens in order to identify antibodies in the array with binding specificity for the antigen. The antigens are often proteins, although they may also be organic chemical compounds, carbohydrates, nucleic acids, and the like. They may be isolated or semi-isolated, recombinant or naturally occurring. The amount of antigen used can vary from about 1 ng to about 1 µg. ~~Because of the small size of the array smaller amounts of protein are preferred.~~ The antigen is left in contact ~~with the array for an~~ amount of time sufficient for antibody:antigen complexes to form, should one of the antibodies in the array be specific for the antigen. The amount of time sufficient for this purpose will range from 5 minutes to 24 hours, and will generally be from 2 to 8 hours.

deleted

One antigen of particular interest in the practice of the invention is a peptide encoded by an Expressed Sequence Tag (or EST fragment). EST fragments are relatively short cDNA sequences that have been randomly generated and sequenced, generally as part of an ongoing effort to map an entire genome (Adams, et al, Science 252(5013):1651-1656, 1991). Large numbers of these sequences are available in public databases. The identity of the proteins encoded by the vast majority of these sequences is unknown.

Techniques are available in the art by which cells can be genetically engineered to express the peptide encoded by a given EST fragment. The methods of the

invention can then be used to identify antibodies specific for the peptide. These antibodies are then useful as reagents that can be employed in purification and identification of the full-length protein, and in other experimental procedures designed to elucidate the protein's location and function.

Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for EST fragments. Prokaryotes most frequently are represented by various strains of E. coli. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119, and the like; suitable phage or bacteriophage vectors may include λgt10, λgt11, and the like; and suitable virus vectors may include pMAM-neo, PKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as E. coli and those from genera such as Bacillus, Streptomyces, Pseudomonas, Salmonella, Serratia, and the like. However, under such conditions, the polypeptide will not be glycosylated. The prokaryotic host selected for use herein must be compatible with the replicon and control sequences in the expression plasmid.

To express an EST fragment in a prokaryotic cell, it

is necessary to operably link the gene sequence to a functional prokaryotic promoter such as the T7 promoter or RSC promoter. Such promoters may be either constitutive or, more preferably, regulatable (i.e., inducible or derepressible). Examples of constitutive promoters include the int promoter of bacteriophage λ, the bla promoter of the β-lactamase gene sequence of pBR322, the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage (P_L and P_R), the trp, reca, lacZ, LacI, and gal promoters of E. coli, the α-amylase (Ulmanen Ett at., J. Bacteriol. 162:176-182, 1985) and the sigma-28-specific promoters of B. subtilis (Gilman et al., Gene sequence 32:11-20(1984)), the promoters of the bacteriophages of Bacillus (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), Streptomyces promoters (Ward et at., Mol. Gen. Genet. 203:468-478, 1986), and the like. Exemplary prokaryotic promoters are reviewed by Glick (J. Ind. Microtiot. 1:277-282, 1987); Cenatiempo (Biochimie 68:505-516, 1986); and Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome binding site upstream of the gene sequence-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold et at. (Ann. Rev. Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene.

Host cells which may be used in the expression

systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the peptide of interest. Suitable hosts may often include eukaryotic cells. Preferred eukaryotic hosts include, for example, yeast, fungi, insect cells, and mammalian cells either in vivo, or in tissue culture. Mammalian cells which may be useful as hosts include HeLa cells, cells of fibroblast origin such as VERO, 3T3 or CHOK1, HEK 293 cells or cells of lymphoid origin (such as 32D cells) and their derivatives. Preferred mammalian host cells include SP2/0 and JS58L, as well as neuroblastoma cell lines such as IMR 332 and PC12 which may provide better capacities for correct post-translational processing.

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, nopaline synthase promoter and polyadenylation signal sequences, and the like. Another preferred host is an insect cell, for example the Drosophila larvae. Using insect cells as hosts, the Drosophila alcohol dehydrogenase promoter can be used. Rubin, Science 240:1453-1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of peptide encoded by an EST fragment in insects cells (Jasny, Science 238:1653, 1987); Miller et al., In: Genetic Engineering (1986), Setlow, J.K., et al., eds., Plenum, Vol. 8, pp. 277-297).

Any of a series of yeast gene sequence expression systems can be utilized which incorporate promoter and termination elements from the actively expressed gene sequences coding for glycolytic enzymes which are

produced in large quantities when yeast are grown in media rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. Yeast provides substantial advantages in that it can also carry out posttranslational peptide modifications. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian gene sequence products and secretes peptides bearing leader sequences (i.e., pre-peptides). For a mammalian host, several possible vector systems are available for the expression of and EST fragment.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

Expression of an EST fragment in eukaryotic

hosts involves the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer et al., J. Mol. Appl. Gen. 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, Cell 31:355-365, 1982); the SV40 early promoter (Benoist et al., Nature (London) 290:304-310, 1981); the yeast gal4 gene sequence promoter (Johnston et al., Proc. Natl. Acad. Sci. (USA) 79:6971-6975, 1982); Silver et al., Proc. Natl. Acad. Sci. (USA) 81:5951-5955, 1984), the CMV promoter, the EF-1 promoter, and the like.

An EST fragment and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA (or RNA) molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule (a plasmid). Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent or stable expression may occur through the integration of the introduced DNA sequence into the host chromosome.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics,

or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by cotransfection. Additional elements may also be needed for optimal synthesis of single chain binding protein mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, Mol. Cell. Bio. 3:280, 1983.

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Suitable prokaryotic vectors include plasmids such as those capable of replication in E. coil (for example, pBR322, ColE1, pSC101, PACYC 184, itVX, pRSET, pBAD (Invitrogen, Carlsbad, CA) and the like). Such plasmids are disclosed by Sambrook (cf. "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, (1989)). Bacillus plasmids include pC194, pC221, pT127, and the like, and are disclosed by Gryczan (In: The

Molecular Biology of the Bacilli, Academic Press, NY (1982), pp. 307-329). Suitable Streptomyces plasmids include plJlOl (Kendall et al., J. Bacteriol. 169:4177-4183,1987), and streptomyces bacteriophages such as ϕC31 (Chater et al., In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary (1986), pp. 45-54). Pseudomonas plasmids are reviewed by John et al. (Rev. Infect. Dis. 8:693-704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729-742, 1978).

Suitable eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, pCDN3.1 (Invitrogen) and the like, or their derivatives. Such plasmids are well known in the art (Botstein et al., Miami Wntr. Symp. 19:265-274, 1982); Broach, In: "The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470 (1981); Broach, Cell 28:203-204, 1982); B(Dilon et at., J. Clin. Hematol. Oncol. 10:39-48, 1980); Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608 (1980).

Once antibody:antigen complexes have been formed and unbound antigen washed away under suitable conditions, the antibody:antigen complexes can be detected using one of several techniques known in the art. Suitable washing conditions are known to those skilled in the art (see, for example, Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995). Examplary washing conditions are shown in the examples below.

For detection in the case of recombinant antigens, expression vectors can be used that form chimeric

peptides comprising the antigen and an epitope tag such as one from hemagglutinin (HA) (Pati, Gene 114(2):285-288, 1992) or glutathione-S-transferase. The epitope tagged antigen can be detected using an antibody specific for the tag sequence. This antibody may be itself detectably labeled or can be detected with a third detectably-labeled antibody. Alternatively, the antigen can be complexed with biotin and detected using detectably-labeled avidin or strepavidin. In a preferred method, the antigen itself is detectably labeled, such as with a fluorescent dye compound.

Substances suitable for detectably labeling proteins include fluorescent dyes such as fluorescein isothiocyanate (FITC), fluorescein, rhodamine, tetramethyl-rhodamine-5-(and 6)-isothiocyanate (TRITC), Texas red, cyanine dyes and the like; and enzymes that react with colorometric substrates such as horseradish peroxidase. The use of fluorescent dyes is generally preferred in the practice of the invention as they can be detected at very low amounts. Furthermore, in the case where multiple antigens are reacted with a single array, each antigen can be labeled with a distinct fluorescent compound for simultaneous detection. Labeled spots on the array are detected using a fluorimeter, the presence of a signal indicating an antigen bound to a specific antibody.

The formation of antibody:antigen complexes can be performed under a variety of conditions to identify antibodies with varying binding characteristics. Antigen-containing reaction solutions can contain varying degrees of salt or be conducted at varying pH levels. In addition, the binding reaction can be carried out at

varying temperatures. Each set of conditions will identify antibodies with different affinity for the antigen. For example, antibodies that bind at pH 2 may have utility under highly acidic conditions such as those that exist in the stomach. Similarly, antibodies that bind at temperatures near boiling may be useful in studying thermophilic organisms. In general pH conditions will range from 2 - 10, temperatures from 0° C - 100°C and salt conditions from 1 μM to 5 M (in the case of NaCl).

Affinity constants are a measure of the interaction between a particular ligand and its cognate receptor. The "binding affinity" or the measure of the strength of association between a particular antibody:antigen interaction is generally measured by affinity constants for the equilibrium concentrations of associated and dissociated configurations of the antibody and its antigen. Preferably the binding of the antigen should occur at an affinity of about $k_a = 10^{-6}$M or greater to be useful for the present invention, with greater than about 10^{-7}M being more preferable, and most preferably between about 10^{-8}M and about 10^{-11}M.

In another embodiment of the invention, microarrays of uncharacterized antibodies are used to compare the protein expression profiles of cells. For example, comparisons can be made between a population of cells from one tissue, such as arterial endothelial cells, and a second tissue, such as venous endothelial cells. Comparisons can be made between normal cells and cells from the same tissue type that originate from an individual with a pathogenic disorder. For example, comparisons can be made between normal cells and cancer

cells. Comparisons can additionally be made between cells in a resting state and cells in an activated state, for example, resting T-cells and activated T-cells.

In another example, the disclosed arrays are useful for evaluating the expression of proteins by pathogens, such as, for example, bacteria, parasites, viruses, and the like. A solution (such as a lysate) made from the pathogen which represents all proteins expressed by the pathogen can be used to contact an antibody array to identify antibodies recognizing pathogen-expressed proteins. These antibodies have utility as diagnostic agents as well as potential therapeutics.

Cellular lysates can be used as "antigens" as described above and reacted with two identical microarrays. Antibodies reactive in one array but not the other would indicate the presence of a differentially expressed protein. This antibody is then useful for the subsequent isolation and identification of those proteins that are different in two populations of cells. In the case of normal and cancer cells, for example, one may be able to identify proteins [solely] expressed in the cancer cell that contribute to its malignant state.

In a further aspect of the invention, microarrays can be composed of previously characterized antibodies. These microarrays have a variety of uses, one of which is cell profiling. For example, an array can be composed of antibodies that recognize a set of antigens known to be present in activated T-cells but not in resting T-cells. A population of T-cells can then be lysed and the lysate contacted with the array to determine if the population has the profile of activated or resting T-cells.

In another aspect of the invention microarrays are employed to characterize protein expression patterns using nucleic acid samples. Briefly, nucleic acid molecules from a whole cell or tissue are applied to a solid support using a microarray format. The arrayed nucleic acid samples are then contacted with a nucleic acid probe specific for a gene encoding a known protein. The probe solution is left in contact with the array for an amount of time sufficient to allow sample:probe complexes to form, then the unbound probe is washed away under suitable conditions (see, for example, Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995 and the examples below). Bound probe is detected at one or more nucleic acid sample spots using one of a variety of detection methods.

This aspect of the invention has a variety of uses. For example, the microarray can be constructed from nucleic acid samples isolated from a single tissue type but from a large number of species, with each spot representing a particular species. Thus in a single assay format one can determine the evolutionary development of the protein represented by the probe. Similarly, the microarray can be constructed of multiple tissues types from a single species, or from different developmental stages of a single species (or multiple species) thus simply and efficiently determining tissue expression of the protein represented by the probe. For example, a microarray can be constructed with arrayed samples representing all the developmental stages of Drosophila, a well known organism the study of which has led to a greater understanding of mammalian physiology and development.

The nucleic acid sample can be messenger ribonucleic acid (mRNA) or can be complementary dexoyribonucleic acid (cDNA), including EST fragments. Methods for extracting and isolating nucleic acids from cells are well known in the art (for example phenol extraction/ethanol precipitation, ammonium acetate precipitation, cesium chloride gradients, and the like), as are methods for generating cDNA (see, for example, "Molecular Cloning: A Laboratory Manual," second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989; and Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995, both of which are incorporated by reference herein). Microarrays of these nucleic acids are created using the methods described above. Techniques for coupling nucleic acids to solid supports used to construct microarrays are well known in the art, including the poly-L-lysine and phenylboronic acid methods described in the Examples below.

The nucleic acid probes used in the invention methods can be designed based on the sequence of a gene encoding a known protein or can be an EST fragment, as described above. One skilled in the art can readily design such probes based on the known sequence using methods of computer alignment and sequence analysis known in the art (e.g. "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989; Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995). The probe can comprise any number of nucleotides but will preferably be not fewer than 10 nucleotide and preferably not more than about 300 nucleotides in length.

The probes of the invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescent label., and the like. After hybridization, the probes may be detected using known methods. Preferred labels are fluorescent labels, as described above.

The nucleic acid probes of the present invention include RNA as well as DNA probes and nucleic acids modified in the sugar, phosphate or even the base portion as long as the probe still retains the ability to specifically hybridize under conditions as disclosed herein. Such probes are generated using techniques known in the art.

The term "hybridize" as used herein refers to a method of interacting a nucleic acid sequence with a DNA or RNA molecule in solution or on a solid support, such as cellulose or nitrocellulose. If a nucleic acid sequence binds to the DNA or RNA molecule with sufficiently high affinity, it is said to "hybridize" to the DNA or RNA molecule. The strength of the interaction between the probing sequence and its target can be assessed by varying the stringency of the hybridization conditions. Various low to high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. Stringency is controlled by varying salt or denaturant concentrations. Examples of hybridization conditions are shown in the Examples below. Those skilled in the art readily recognize how such conditions can be varied to vary specificity and selectivity. For example, under highly stringent hybridization conditions only highly complementary

nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having even one or two mismatches out of 20 contiguous nucleotides.

In a further aspect of the invention, microarrays can be composed of randomly generated polynucleotides (DNA or RNA) and contacted with proteins to identify unique binding pairs. Polynucleotides are now known to bind to proteins and may have potential as diagnostics and therapeutics (see, for example, Allen, et al, Virology 209(2):327-336, 1995; Binkley, et al, Nucleic Acids Res. 23(16):3198-3205, 1995). Polynucleotides can be evaluated in very large numbers using the methods disclosed herein thus increasing the likelihood of identifying a useful binder.

The invention will now be described in greater detail by reference to the following non-limiting examples.

Examples

The following procedures are conducted at room temperature and using double distilled water unless otherwise noted. These methods are applicable to arrays of polypeptides or polynucleic acids.

Glass slides are prepared as follows: NaOH (50 g) is dissolved in 150 ml of double distilled water (ddH_2O), then 200 ml of 95% EtOH is added while stirring. If the solution becomes cloudy, ddH_2O is added until it becomes

clear. Approximately 30 glass slides (Gold Seal, Cat. No. 3010) are soaked in the NaOH/EtOH solution for 2 hours, shaking. The slides are then rinsed three times with ddH_2O. The slides are next soaked in a poly-L-lysine solution (70 ml poly-L-lysine (Sigma Cat. No. 8920) to 280 ddH_2O) for 1 hour. Excess liquid is removed by spinning in the slides in a rack on a microtiter plate carrier at 500 rpm. The slides are dried at 40° C for 5 minutes, then stored in a closed box for at least 2 weeks prior to use.

A cDNA microarray is prepared as follows: Total mRNA from is isolated from tissue (for example, nerve cells) of a variety of species representative of different classes of organisms such as Drosophila, nematode, salmon, clam, chicken, mouse, dog, goat, spider monkey, chimpanzee, human, and the like, by the FastTrac method (Stratagene, La Jolla, CA) or other common methods. mRNA is also obtained from a variety of unicellular organisms such as E. coli, yeast, B. subtilis, mycoplasma and the like. Eurokaryotic mRNA is enriched from total RNA using oligo(dT) cellulose (Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed. 1995, pgs 4-11 - 4-12). Equivalent amounts (for example, 1 µg) of mRNA from each source is placed in a separate well of one or more 96 well microtiter plates and precipitated with cold EtOH. The precipitate is rinsed with 70% EtOH and allowed to dry.

The dried mRNA is resuspended in 3x SSC (sodium chloride/sodium citrate - 20X solution is 3 M NaCl (175g/L0 0.3 M trisodium citrate $2H_2O$ (88g/L adjusted to pH 7.0 with 1 M HCl) then spotted onto a previously prepared glass slide using an array device (for example,

Shalon and Brown (WO 95/35505, published 12/28/95)). The prepared array can be kept for a long period of time before probing, however, if the slides are to be kept for long periods of time, stability is increased by converting each mRNA sample into cDNA using techniques known in the art, such as PCR.

The array is rehydrated by suspending the slide over a dish of ddH_2O (50°C) for approximately one minute. The slide is quickly (approximately 3 seconds) dried by placing it on a surface heated to 100° C (mRNA side up). The mRNA is crosslinked to the poly-L-lysine coating of the slide using ultraviolet radiation using a Stratalinker™ UV device according to the manufacturer's instructions (Stratagene) set at 60 milliJoules.

The slides are next soaked in a solution of 5 grams of succinic anhydride (Aldrich Cat. No. 23,969-0) dissolved in 315 ml of n-methyl-pyrrilidinone (Aldrich Cat. No. 32,863-4) plus 35 mls of 0.2 M Na Borate (brought to pH 8.0 with NaOH) for 15 minutes with shaking. The slide is then transferred to a 95° C water bath for 2 minutes followed by 95% EtOH for 1 minute. Excess liquid is removed from the slides by spinning a rack of slides on a microtiter plate carrier at 500 rpm.

A probe sequence of a known protein (for example, human nerve growth factor, GeneBank Accession NO. EO3589) is labeled using standard protocols, for example by using a CyDye Nick Translation kit (Amersham). The labeled probe (approximately 1 µg/ml) is resuspended in 4X SSC (10 µl) to which is added 0.2 µl 10% sodium dodecyl sulfate (SDS). The probe is boiled for 2 minutes, then cooled for 10 seconds and transferred to the array by

pipette. The array is covered by a 22mm x 22mm cover slip, and the slide is placed in a humid hybridization chamber and submerged into a hot water bath ($\geq$ 75°C).

The slide is left in the bath for 10 - 24 hours, then the cover slip is removed and the slide rinsed in 0.2X SSC with 0.1% SDS several times. Excess wash buffer is removed by centrifugation on a microtiter plate carrier is described above. The slide is scanned using a spectrofluorometer, such as the ScanArray 3000 (General Scanning Inc., Watertown, MA). For probes labeled with Cy5, for example, fluorescence is measured at 670 nm.

Localization of spots on the array to which the probe hybridizes indicates that the species represented by the spot expresses a protein similar or identical to the probe protein.

The procedure outlined below is an alternative method for binding arrayed molecules to a solid support, using an SA(OCH_2CN)-X-NHS linkage (see, for example, US 5,594,111, issued 1/14/97; 5,648,470, issued 5/15/97; US 5,623,055, issued 4/22/97; all of which are incorporated by reference herein).

Glass slides (Fisher Catalog No. 12-544-4) are soaked in an acid bath (1 hour in 0.1 M HCl), then washed with water and dried at room temperature. The slides should not be aggressively dried, such as in an oven. The slides are next soaked in a silane solution overnight at room temperature (5% APTES (3-aminopropyltriethoxysilane, Aldrich 28,177-8), 0.3% DIEA (Sigma) v/v in EtOH). The slides may be sonicated for 10 - 15 minutes right after being placed in the APTES solution.

The slides are rinsed with isopropyl alcohol, then sonicated in isopropyl alcohol for several minutes. Sonication should remove any white silane residue on the slides. If the residue remains, the slides should be discarded. After sonication, the slides are left to cure/dry for at least 24 hours before use.

The cured slides are next soaked in a linker solution overnight at room temperature. The linker solution is made by dissolving 115 mg of 9Y SA(OCH_2CN)-X-COOH (Prolix, Bothell, WA) in 1 ml dimethylformamide (DMF) plus 60 µl DIEA, then adding 60 mg TSTU (Sigma) and leaving for 15 minutes at room temperature. This stock is diluted in 270 ml of isopropyl alcohol plus 270 µl DIEA before using.

The slides are removed from the linker solution and soaked in 1 M NH_2OH, 1mM EDTA, 0.1 M $NaHCO_3$ (pH 10) for 4 hours at room temperature. This solution is removed, the slides are extensively washed with water then let air dry at room temperature. The slides can be stored at room temperature away from light before using to make arrays.

While the foregoing has been presented with reference to particular embodiments of the invention, it will be appreciated by those skilled in the art that changes in these embodiments may be made without departing from the principles and spirit of the invention, the scope of which is defined by the appended claims.

That which is claimed is:

1. A method of identifying antibodies having binding affinity for a specific target antigen, said method comprising:

(a) contacting an array of unidentified single chain antibodies on a solid surface with at least one target antigen; and

(b) identifying the antibodies to which the target antigen binds.

2. A method according to claim 1 wherein the target antigen is a protein, a peptide, an intact cell or a cell extract.

3. A method according to claim 2 wherein the peptide is encoded by an EST fragment.

4. A method according to claim 1 further comprising varying the binding conditions under which said contacting is carried out, thereby enabling the determination of the effect of said conditions on the binding affinity of said antibodies for said specific target antigen.

5. A method according to claim 4 wherein said binding conditions comprise varying pH, temperature, salt concentration, or duration

6. A method according to claim 1 wherein the binding affinity of said antibody for said antigen is determined by iterative washing of said solid surface with a suitable diluent and detecting when antigen is no longer released therefrom.

7. A method to determine gene expression at the protein level, said method comprising:

(a) contacting an array of unidentified single chain antibodies on a solid surface with said protein; and

(b) identifying the antibodies to which said protein binds.

8. A method of comparing two populations of cells, said method comprising:

(a) contacting a first array of unidentified antibodies on a solid surface with a cell extract of a first set of cells, generating a first binding pattern;

(b) contacting a second array of unidentified antibodies on a solid surface with a cell extract of a second set of cells, generating a second binding pattern; and

(c) comparing the binding pattern of the first cell extract with the binding pattern of the second cell extract.

9. A method according to claim 8 wherein the first cell extract is made from normal cells and the second cell extract is made from abnormal cells.

10. A method according to claim 9 wherein the abnormal cells are cancer cells.

11. A method according to claim 9 wherein the first cell extract is made from normal cells in a resting state and the second cell extract is made from normal cells in a stimulated state.

12. A method according to claim 8 wherein said first array and said second array are the same.

13. A method according to claim 12 wherein the cell extract from said first set of cells and the cell extract from said second set of cells is the same, but the cell extract from said first set of cells is labeled with a different label than is the cell extract from said second set of cells.

14. A method for determining the effect of varying binding conditions on the binding affinity of antibodies having a specific target antigen, said method comprising:

(a) contacting an array of single chain antibodies on a solid surface with at least one target antigen under a first set of binding conditions;

(b) contacting said array of single chain antibodies on a solid surface with said target antigen under a second set of binding conditions; and

(c) identifying the difference in binding affinity of said antibodies to said target antigen as a function of the different conditions employed in steps (a) and (b).

15. A method according to claim 14 wherein said binding conditions comprise varying pH, temperature, salt concentration, or duration

16. A method for characterizing a cell, based on the pattern of protein expression produced thereby, said method comprising:

(a) contacting an array of single chain antibodies on a solid surface with an extract of proteins produced by said cell; and

(b) identifying the profile of antibodies to which members of said extract binds.

17. A method of comparing protein expression patterns comprising:

(a) contacting a microarray of nucleic acid samples derived from different sources with one or more nucleic acid probes and

(b) identifying the sample or samples to which the probe binds.

18. A method of claim 17 wherein the microarray comprises nucleic acid samples derived from a single tissue type but from different species.

19. A method of claim 17 wherein the microarray comprises nucleic acid samples derived from a single species but from different tissue types.

20. A method of claim 17 wherein the microarray comprises nucleic acid samples derived from the same tissue type at different developmental stages.

21. A method of claim 17 wherein the nucleic acid samples are comprised of mRNA or cDNA.

22. A method of claim 17 wherein the probe is detectably labeled.

23. A method of claim 22 wherein the detectable label is a fluorescent label.